

Available for Licensing

# Nucleic Acid Analysis Software

*Following nucleic acid analysis software is available for licensing from Argonne National Laboratory*

**FASTA\_EXCLUDE:** This software allows generating FASTA alignment file based on existing FASTA alignment database that doesn't contain specified sequences. As an auxiliary output, the program specifically lists sequences that were removed. Input and output file formats: FASTA. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-111

**REMOVE\_COMMON\_GAPS:** This program removes all gaps from FASTA alignment file that are present in all sequences, leaving all sequences aligned by in smaller file (i.e. remove all columns that are completely empty). Input and output file formats: FASTA. Developers/Authors: Sergey Bavykin and Oleg Alferov ANL-SF-06-111b

**TREE\_SELECT\_PROBES:** This program is part of 3 program package that replaces the older probe selection software. The purpose of the package is to generate probes specific for the group of sequences that belong to a given phylogenetic node. This software employs modified probe selection algorithm that improves speed of calculations in comparison with older software. For each node of the input tree, this program selects probes that are positive for all sequences that belong to this node and negative for all that doesn't. For speed, it uses probe database created by build\_hybr\_index program and hybridization table database created by build\_hybrid\_table program. As a result of calculation, the program prints lists for each node from the tree. Input file forms: FASTA for sequence database, internal INDEX for probe database, internal table for hybridization database. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-110

**HYBINFO:** This software converts the two dimensional table of hybridization intensity information read from fluorescent microscope into a list organized by probe sequences, based on the chip map with X-Y coordinated and probe content. The program also provides simulated hybridization prediction for all probes, according to the sequence of test specimen (if known). Input file formats: text list for probe sequences and coordinates, tab-delimited text file for hybridization data table, single line text file for sequence. Output format: CSV text file. Developers/Authors: Sergey Bavykin and Oleg Alferov. ANL-SF-06-112

**95MARGIN V2:** This software is intended to trim the alignment file from left and right removing the portions of sequences that extend beyond alignment kernel limits. Simultaneously this program can remove sequences that look "bad" i.e. too short, too offset, contains multiple N-s, generates excessive gaps in the entire alignment, etc. User controls the program behavior by specifying the percent of sequences to keep, and the algorithm automatically decides how aggressive the trimming should be. Input and output file formats: FASTA. Developers/Authors: Sergey Bavykin and Oleg Alferov. ANL-SF-06-112e

**TREE\_SELECT\_PROBES:** This program is part of a 3 program package that replaces the older probe selection software. The purpose of the package is to generate probes specific for the group of sequences that belong to a given phylogenetic node. This software employs modified probe selection algorithm that improves speed of calculations in comparison with older software. For each node of the input tree, this program selects probes that are positive for all sequences that belong to this node and negative for all that doesn't. For speed it uses probe database created by build\_hybr\_index program and hybridization table database created by build\_hyper\_table program. As a result of calculation, the program prints lists

for each node from the tree. Input file formats: FASTA for sequences database, internal INDEX for probe database, internal table for hybridization database. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-110

**BUILD\_HYBR\_INDEX:** This program is part of a 3 program package that replaces the older probe selection software. The purpose of the package is to generate probes specific for the group of sequences that belong to a given phylogenetic node. This software employs modified probe selection algorithm that improves speed of calculations in comparison with older software. This program creates database of potential probes based on given sequence database, reducing it in the way so it doesn't contain repeats or substrings with meta-nucleotides. Input file format: FASTA. Output file format: internal INDEX file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-110c

**BUILD\_HYBR\_TABLE:** This program is part of a 3-program package that replaces the older probe selection software. The purpose of the package is to generate probes specific for the group of sequences that belong to a given phylogenetic node. This software employs modified probe selection algorithm that improves speed of calculations in comparison with older software. For each node of the input tree, this program selects probes that are positive for all sequences that belong to this node and negative for all that doesn't. For speed, it uses probe database created by build\_hybr\_index program and hybridization table database created by build\_hybr\_table program. As a result of calculation, the program prints lists for each node from the tree. Input file formats: FASTA for sequence database, internal INDEX for probe database, internal table for hybridization database. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-110b

**SNAME\_SUBST V2:** This software substitutes RDP-specific S-numbers of microorganisms with their human-readable equivalents in any given text file, based on FASTA database. Input file formats: original text file, FASTA database. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-111b

**NWK\_ARB\_APPLY:** This software substitutes RDP-specific S-numbers of microorganisms with their human-readable equivalents preceded by phylogenetic information. The phylogenetic information is defined by the input tree file. The program also employs measures to create acronyms instead of full names to save space. Input file formats: original text file. FASTA database for names substitution. ARB tree for representation of phylogenetic information. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-112b

**1MISS:** Searches for the pairs of DNA/RNA probes that differ from each other by one nucleotide, based on the given list of oligonucleotide probes. Input and output file formats: text lists. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-112d

**NWK LIM:** The program modifies the phylogenetic tree in the NWK format removing all the nodes with score below the given number, creating another NWK tree, which is reduced but has valid format, input and output formats: NWK tree. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108

**NWK\_TREE\_SPLIT:** The program divides one tree in two; one is a subtree that starts from the given node, and the other is the tree that is cut at this node. (i.e. two complementary trees that are formed by removing a link at the given code). Input and output file formats. NWK trees. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108d

**REVERSE:** This program splits the file with probes and nodes produce by tree\_select\_probes programs into separate files that contain all probes (with all parameters: quality, coverage, distance, etc. for a specific node. For each node, the separate file is written. Input file format: tree\_select\_probes output probe and node list. Output file format: text files with list of probes with their parameters. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-107

**OVERLAP\_FILTER\_GRAD:** The program performs intelligent filters on a probe list arranged into blocks,

according to minimum Quality (q). coverage (p), distance, hairpins, self-hybridization, minimum number of probes for the node and for the block, etc. Input and output file formats: text files with probe lists. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-107c

**OVERLAP\_COLLECT:** This program reads all data associated with probe sequences and node names, and creates a recognition database that is list of all probe sequences, with list of matching nodes for each probe. Input file format; text file that is mechanically combined from all node and probe files. Output file format: text file, recognition database. Developers/Authors: Sergey Bavykin and Oleg Alferov. ANL-SF-07-107d

**ARB\_TREE\_32:** The purpose of this program is to generate probes specific for the group of sequences that belong to a given phylogenetic node. For each node of the input tree, this program selects probes that are positive for all sequences that belong to this node and negative for all that doesn't. The program uses condensed tree for probe representation to save computer memory. As a result of calculation, the program prints lists for each node from the tree. Input file formats: FASTA for sequence database and ARB tree for phylogenetic organization of nodes. Output file format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046

**VIRT\_HYB:** This program compares the given list of nucleotide probes versus the given nucleotide sequence and prints 1 (true) or 0 (false) for each probe, when it matches or doesn't match the sequence. Input formats: text files for sequence and list of probes. Output format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046f

**NWK\_NUM:** The program assigns numbers to all nodes in the input tree, in the way that the number is displayed by MEGA software. Input and output file formats: NWK tree. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108b

**NWK\_NUM\_STRIP:** This program removes the node numbers from the NWK tree, assigned by nwk\_num program, but leaves the original score values. Input and out file formats: NWK tree. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108c

**NWK\_TREE\_MERGE:** The program takes two phylogenetic trees and insert one into the place of the given leaf of another tree. User must specify the leaf to replace, and the new score number that will be assigned to the root of the insert. Input and output file formats. NWK trees. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108e

**NWK\_NARB\_PCT:** This program writes node names and score into the text list. Input format: NWK tree. Node names matches those created by nwk2arb program tool. Output format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-108f

**OVERLAP\_EQU:** The program examines the list of probes and selects the series of sequences that are continuations of each other, and group such the sequences into blocks. Input and output file formats: text files with list of probe sequences. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-107b

**COMPLEMENT:** This program calculates complementary sequence to the given nucleotide sequence. Input format: text file with list of input sequences separated by newline. Output format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046b

**NWK2ARB:** The program converts the phylogenetic tree from NWK format to ARB format. Input format: NWK file. Output format: ARB file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046c

**PQFILTER:** This program filters the list of probes created by arb\_tree\_32 by specified criteria on cover (P) and quality(Q) parameters. Input format: text file with list of probes. Output format: text file.

Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046d

**REQUALIFY:** This program recalculates the Q (quality) parameter for probe list generated by arb\_tree\_32 program using different (usually larger) FASTA database. Input formats: test file with probe list, FASTA database. Output format: text file. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-07-046e

**REMOVE\_COMMON\_GAPS:** This program removes all gaps from FASTA alignment file that are present in all sequences, leaving all sequences aligned but in smaller file. (i.e., remove all columns that are completely empty.) Input and output file formats: FASTA. Developers/Authors: Sergey Bavykin and Oleg Alferov, ANL-SF-06-111b

**Availability of Software** Above mentioned nucleic acid analysis software is available for licensing from Argonne National Laboratory

References: None

#### CONTACT

**Argonne Technology Development and Commercialization**

Phone: 800-627-2596

E-mail: [partners@anl.gov](mailto:partners@anl.gov)



Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC. October 2016